AD-A107 520 BROWN UNIV PROVIDENCE RI LEFSCHETZ CENTER FOR DYNAMI--ETC F/G 12/2
APPROXIMATION OF PROCESSES AND APPLICATIONS TO CONTROL AND COMM--ETC(U)
JUN 81 H J KUSHNER N00014-76-C-0279
UNCLASSIFIED LCDS-81-18 NL

1 of 1
AD
A107520

END
DATE
FILMED
10 81
DTIC

Lefschetz Center for Dynamical Systems

APPROXIMATION OF PROCESSES AND APPLICATIONS TO

CONTROL AND COMMUNICATION THEORY

by

Harold J. Kushner

June 1981 LCDS ~~Report~~ 81-18

APPROXIMATION OF PROCESSES AND APPLICATIONS TO

CONTROL AND COMMUNICATION THEORY†

Harold J. Kushner

Divisions of Applied Mathematics and Engineering
Lefschetz Center for Dynamical Systems
Brown University
Providence, Rhode Island 02912

June 1981

†Research supported in part by the Air Force Office of Scientific Research under AFOSR 76-3063D, in part by the National Science Foundation under NSF-Eng 77-12946-A02 and in part by the Office of Naval Research under N00014-76-C-0279-P0004.

APPROXIMATION OF PROCESSES AND APPLICATIONS TO CONTROL AND COMMUNICATION THEORY

by

Harold J. Kushner

Abstract and Introduction

Diffusion models are useful and of widespread use in many areas of control and communication theory. The models are frequently used for systems that are not quite diffusions but are, hopefully, close to a diffusion in some sense. E.g., the input noise might be 'wide-band'-- but not 'white-Gaussian.' Many approximation techniques have been developed ~~[1]-[10]~~, under different sets of assumptions. The typical results are of a weak convergence nature. The physical process $x^{\varepsilon}(\cdot)$ is parameterized by the parameter ε, and one tries to show that $\{x^{\varepsilon}(\cdot)\}$ converges weakly to some diffusion $x(\cdot)$ as $\varepsilon \to 0$. The limit process $x(\cdot)$ is then used to study various properties of $x^{\varepsilon}(\cdot)$ for small ε.

If the system involves nonlinear functions of the input noise, or if the noise depends on the system state or if the system dynamics are not smooth, then the approximation problem is harder. Here, we discuss a set of techniques which have proved to be quite useful for a variety of problems in control and communication. The main weak convergence theorems are stated and discussed in Section II. Frequently, in applications, we are concerned with asymptotic properties, as $t \to \infty$ (for small ε), as well as with weak convergence. Such information is not normally provided by the weak convergence theory. In Section II, we also discuss the prob-

lem of approximating functionals on the 'tail' of $x^{\varepsilon}(\cdot)$ for small ε, by such functionals on the 'tail' of $x(\cdot)$ (e.g., approximating the measures of $\{x^{\varepsilon}(t)$, large $t\}$, for small ε, by an invariant measure of $x(\cdot)$. This is particularly useful in problems in communication theory, where the (say) detection system is often supposed to be in operation for a very long time. The proofs of these theorems are given in the references.

In Sections III and IV, we apply these theorems to two systems which are of frequent and basic use in communication theory, but which seem to be quite difficult to analyze without at least some sort of formal diffusion approximation technique. Other applications to control and communication of these and related ideas can be found in [9], [11]-[14]. While there has been widespread application of diffusion approximation methods to some topics in physics and biology, and operations research (e.g., to queueing theory), applications to concrete nonlinear problems in control and communication theory is still in its infancy. There are numerous possibilities for other applications; e.g., to synchronization systems, robustified filters (where some nonlinear function of the observation is used in lieu of the observation), to adaptive modulation systems and filters, etc.

II. CONVERGENCE AND APPROXIMATION THEOREMS

Weak Convergence. Suppose that the system state $x^\varepsilon(\cdot)$ satisfies the differential equation $\dot{x}^\varepsilon = H^\varepsilon(\xi^\varepsilon, x^\varepsilon)$, where $\xi^\varepsilon(\cdot)$ is an input noise process whose bandwidth (BW) goes to ∞ as $\varepsilon \to 0$ (loosely speaking), and $x^\varepsilon(t) \in R^r$, Euclidean r-space. We are first interested in showing that $\{x^\varepsilon(\cdot)\}$ converges weakly in $D^r[0,\infty)$ to some diffusion process $x(\cdot)$, defined by (2.1), where we assume that the martingale problem in $D^r[0,\infty)$ associated with (2.1) has a unique solution for each initial condition $x = x(0)$.

$$dx = \alpha(x)dt + \sigma(x)dB \quad , \quad B(\cdot) = \text{standard Wiener process} \tag{2.1}$$

It is possible to treat problems where the limits are jump-diffusions as in [7]-[8], but since the applications in Sections III and IV all have diffusion limits, the basic theorems here are specialized to this case.

Define the differential generator

$$A = \sum_i \alpha_i(x)\ \partial/\partial x_i + \frac{1}{2} \sum_{i,j} a_{ij}(x)\ \partial^2/\partial x_i \partial x_j, \text{ where } a(x) = \{a_{ij}(x)\} = \sigma(x)\sigma'(x)\ .$$

For each $N > 0$, let $b_N(\cdot)$ denote a continuously differentiable function satisfying $b_N(x) = 1$ for $x \in S_N = \{x : |x| \le N\}$, $0 \le b_N(x) \le 1$, and $b_N(x) = 0$ for $x \notin S_{N+1}$. Let A^N denote the differential generator of a diffusion $x^N(\cdot)$ with coefficients $\alpha^N(\cdot)$, $\sigma^N(\cdot)$ equal to $\alpha(\cdot)$, $\sigma(\cdot)$, resp. in S_N. Define the truncated process $x^{\varepsilon,N}(\cdot)$ by $\dot{x}^{\varepsilon,N} = H^\varepsilon(\xi^\varepsilon, x^{\varepsilon,N})b_N(x^{\varepsilon,N})$. Then, if $x^{\varepsilon,N}(\cdot) \to x^N(\cdot)$ weakly in $D^r[0,\infty)$, as $\varepsilon \to 0$ for each N, we have [8] $x^\varepsilon(\cdot) \to x(\cdot)$ weakly also. The truncation is used as a technical device only, in that allowing us to work with bounded processes, it simplifies the proofs and the calculations in the

applications of the theorem, although it makes for a slightly more complicated theorem statement.

Let $\hat{C}_0$ denote the space of real valued continuous functions on R^r with compact support, and $\hat{C}_0^{\alpha,\beta}$ the subspace whose mixed α partial t-derivatives and β partial x-derivatives are continuous. Let $\{F_t^\varepsilon\}$ be a non-decreasing sequence of σ-algebras with F_t^ε measuring $\{\xi^\varepsilon(s),\ s \le t\}$. Let H denote the class of real valued (progressively) measurable (ω,t) functions such that if $g(\cdot) \in H$, then $\sup_t E|g(t)| < \infty$, $E|g(t+\delta) - g(t)| \to 0$ as $\delta \to 0$, and $g(t)$ is F_t^ε measurable. Let E_t^ε denote expectation conditioned on F_t^ε. Following [4], we say $\text{p-lim}_{\delta\to 0} f^\delta(\cdot) = 0$ if $\sup_{t\ge 0,\delta>0} E|f^\delta(t)| < \infty$ and $E|f^\delta(t)| \to 0$ as $\delta \to 0$, for each t. Define the operator $\hat{A}^\varepsilon$ with domain $\mathcal{D}(\hat{A}^\varepsilon)$ as follows: $g \in \mathcal{D}(\hat{A}^\varepsilon)$ and $\hat{A}^\varepsilon g(\cdot) = q(\cdot)$ if $g(\cdot)$ and $q(\cdot)$ are in H and $\text{p-lim}_{\delta\to 0} \left[\left(E_t^\varepsilon g(\cdot+\delta) - g(\cdot)\right)/\delta - q(\cdot)\right] = 0$. Thus, $\hat{A}^\varepsilon$ resembles an infinitesimal operator. The following theorem is a specialization of a result in [8], which in turn is a development of ideas of Kurtz [4].

Theorem 1. Let $\alpha(\cdot)$, $\sigma(\cdot)$ be continuous, and assume the above uniqueness condition on the solution of (2.1). Fix N. For each $f(\cdot) \in \hat{C}_0^{2,3}$, let there be a sequence $\{f^{\varepsilon,N}(\cdot)\} \in H$ such that

$$\begin{aligned} &\text{p-lim}_{\varepsilon\to 0} \left|f^{\varepsilon,N}(\cdot) - f\left(x^{\varepsilon,N}(\cdot),\cdot\right)\right| = 0 \\ &\text{p-lim}_{\varepsilon\to 0} \left|\hat{A}^\varepsilon f^{\varepsilon,N}(\cdot) - \left(A^N + \frac{\partial}{\partial t}\right) f\left(x^{\varepsilon,N}(\cdot),\cdot\right)\right| = 0 \ . \end{aligned} \tag{2.2}$$

Then, if $\{x^{\varepsilon,N}(\cdot)\}$ is tight in $D^r[0,\infty)$ for each N, and $x^\varepsilon(0) \to x_0$ weakly, $\{x^\varepsilon(\cdot)\}$ converges weakly to $x(\cdot)$ (with $x(0) = x_0$) as $\varepsilon \to 0$.

<u>Remark</u>. In the examples, it is shown how to get $\{f^{\varepsilon,N}(\cdot)\}$ in typical cases; the method is also discussed in [3], [7], [9], [11]-[14]. We use the form $f^{\varepsilon,N}(t) = f(x^{\varepsilon,N}(t),t) = \sum_{i=0 \text{ or } 1}^{2} f_i^{\varepsilon,N}(t)$, where a convenient method for getting the perturbations $f_i^{\varepsilon,N}$ is discussed in Sections III and IV. In the applications which we have in mind, it is convenient to prove tightness via Theorem 2 below. Let $\hat{C}_0$ denote the space of real valued continuous functions on R^r, with compact support.

<u>Theorem</u> 2 [8]. <u>Fix</u> N. <u>For each</u> $f(\cdot)$ <u>in a dense set</u> $\mathcal{D}_1 \subset \hat{C}_0$ <u>which contains the square of each function in it, let there be a sequence</u> $\{f^{\varepsilon,N}(\cdot)\}$ <u>in</u> H <u>such that</u> $f^{\varepsilon,N}(\cdot) \in \mathcal{D}(\hat{A}^{\varepsilon})$, <u>and for each real</u> $T > 0$, <u>let there be a random variable</u> $M_T^{\varepsilon,N}(f)$ <u>such that</u>

$$P\{\sup_{t\le T} |f^{\varepsilon,N}(t) - f(x^{\varepsilon,N}(t))| \ge \alpha\} \to 0 \quad \text{as} \quad \varepsilon \to 0 \quad , \quad \text{each} \quad \alpha > 0 \ ,$$

$$\sup_{t\le T} |\hat{A}^{\varepsilon} f^{\varepsilon,N}(t)| \le M_T^{\varepsilon,N}(f) \tag{2.3}$$

$$\sup_{\varepsilon>0} P\{M_T^{\varepsilon,N}(f) \ge K\} \to 0 \quad \text{as} \quad K \to \infty \ .$$

Then $\{f(x^{\varepsilon,N}(\cdot))\}$ <u>is tight in</u> $D^1[0,\infty)$ <u>for</u> $f(\cdot) \in \mathcal{D}_1$, <u>and</u> $\{x^{\varepsilon,N}(\cdot)\}$ <u>is tight in</u> $D^r[0,\infty)$.

<u>Remark</u>. Typically, $\mathcal{D}_1 = \hat{C}_0^{2,3}$, but with the time dependence suppressed, and the $f^{\varepsilon,N}(\cdot)$ are constructed as they are for Theorem 1. The conditions are not particularly restrictive, as will be seen in the examples below, or as can be seen from the applications in the references.

<u>Discrete Parameter</u>. Theorem 1 simplifies slightly in the discrete parameter case. Let $\{\tau_\varepsilon\}$ denote a sequence of positive numbers tending to zero as $\varepsilon \to 0$, and suppose that the random sequence $\{x_n^\varepsilon\}$ satisfies $x_{n+1}^\varepsilon = x_n^\varepsilon + h^\varepsilon(\xi_n^\varepsilon, x_n^\varepsilon)$, for some suitable measurable $h^\varepsilon(\cdot.\cdot)$ and random

sequence $\{\xi_n^\varepsilon\}$. Define $\{x_n^{\varepsilon,N}\}$ by $x_{n+1}^{\varepsilon,N} = x_n^{\varepsilon,N} + h^\varepsilon(\xi_n^\varepsilon, x_n^{\varepsilon,N}) b_N(x_n^{\varepsilon,N})$, and define $x^\varepsilon(\cdot)$, $x^{\varepsilon,N}(\cdot)$ by $x^\varepsilon(t) = x_n^\varepsilon$, $x^{\varepsilon,N}(t) = x_n^{\varepsilon,N}$ on $[n\tau_\varepsilon, n\tau_\varepsilon + \tau_\varepsilon)$. Write E_n^ε for $E_{n\tau_\varepsilon}^\varepsilon$. Then for $f(\cdot) \in H$, we now define $\hat{A}^\varepsilon f(n\tau_\varepsilon) = [E_n^\varepsilon f(n\tau_\varepsilon + \tau_\varepsilon) - f(n\tau_\varepsilon)]/\tau_\varepsilon$. In the application of Section 4, $\tau_\varepsilon = \varepsilon$. Theorem 1 can be now rewritten as:

Theorem 3 [8]. Assume the conditions of Theorem 1, but where (2.4) replaces (2.2), and the $f^{\varepsilon,N}(\cdot)$ are constant on the $[n\tau_\varepsilon, n\tau_\varepsilon + \tau_\varepsilon)$ intervals.

$$\sup_{n,\varepsilon} E|f^{\varepsilon,N}(n\tau_\varepsilon)| < \infty , \quad \sup_{n,\varepsilon} E|\hat{A}^\varepsilon f^{\varepsilon,N}(n\tau_\varepsilon)| < \infty ,$$

$$E|f^{\varepsilon,N}(n\tau_\varepsilon) - f(x^{\varepsilon,N}(n\tau_\varepsilon), n\tau_\varepsilon)| \to 0 \tag{2.4}$$

$$E|\hat{A}^\varepsilon f^{\varepsilon,N}(n\tau_\varepsilon) - \left(\frac{\partial}{\partial t} + A^N\right) f(x^{\varepsilon,N}(n\tau_\varepsilon), n\tau_\varepsilon)| \to 0 ,$$

for each t and $n\tau_\varepsilon \to t$ as $\varepsilon \to 0$. Then the conclusions of Theorem 1 continue to hold. (There is also the obvious discrete parameter simplification of Theorem 2.)

Approximation of Asymptotic and Invariant Measures. Theorems 1 to 3 yield weak convergence of $\{x^\varepsilon(\cdot)\}$ to $x(\cdot)$ in $D^r[0,\infty)$. Frequently in applications we are more interested in the measures of $x^\varepsilon(t)$ for large t and small ε. Suppose that $x(\cdot)$ has a unique invariant measure $\mu(\cdot)$. In practice this measure is often used as an approximation to the 'asymptotic' measures of $x^\varepsilon(t)$. In particular, it is of interest to show at least that the measure of $x^\varepsilon(t)$ converges in some sense to $\mu(\cdot)$ as $t \to \infty$, $\varepsilon \to 0$ in any way at all. Weak convergence does not yield such information directly. But weak convergence together with an 'averaged Liapunov function' technique and some assumptions on $x(\cdot)$ can

yield the desired result [15]. Here, we cite some results which will be used in the examples. The basic assumptions are (A1)-(A4), and Theorem 4 is the basic convergence theorem. Conditions guaranteeing (A4) will be given after the statement of Theorem 4.

Theorem 4 requires:

A1. $x(\cdot)$ is a Feller diffusion process with continuous coefficients and a unique weak sense solution on $[0,\infty)$ for each $x(0) = x$.

A2. $x(\cdot)$ has a unique invariant measure $\mu(\cdot)$ and $P_x\{x(t) \in \cdot\}$ converges weakly to $\mu(\cdot)$ for each x, as $t \to \infty$. The convergence is uniform in compact x-sets; i.e., for each $f(\cdot)$ in the space of continuous bounded functions on R^r, $E_x f(x(t)) \to E_\mu f(x(0))$ uniformly in x in compact sets, as $t \to \infty$.

A3. $x^\varepsilon(\cdot) \to x(\cdot)$ (initial condition $x(0)$) weakly if $x^\varepsilon(0) \to x(0)$ weakly in $D^r[0,\infty)$.

A4. There is an $\varepsilon_0 > 0$ such that $\{x^\varepsilon(t),\ 0 < \varepsilon \leq \varepsilon_0,\ t \geq 0\}$ is tight.

Theorem 4 [15]. Assume (A1) - (A4). Then for each integer m, real $T < \infty$, and bounded continuous $f(\cdot)$ and $\delta > 0$, there are $t_0(f,\delta) < \infty$ and $\varepsilon_0(f,\delta) > 0$ such that for $t \geq t_0(f,\delta)$ and $\varepsilon \leq \varepsilon_0(f,\delta)$, and any sequence $\{x^\varepsilon(\cdot)\}$ which converges weakly to $x(\cdot)$ and $T \geq \Delta_i \geq 0$, $i = 1, \ldots, m$,

$$\left|Ef\left(x^\varepsilon(t+\Delta_i),\ i \leq m\right) - E_\mu f\left(x(\Delta_i),\ i \leq m\right)\right| < \delta$$

Let $\left(x^\varepsilon(\cdot), \xi^\varepsilon(\cdot)\right)$ be Markov and have an invariant measure $\nu^\varepsilon(\cdot)$ with x-marginal $\mu^\varepsilon(\cdot)$. Replace (A4) by: There is a sequence T_ε (which

<u>might tend to</u> ∞ <u>as</u> $\varepsilon \to 0$, <u>and</u> T_ε <u>can depend on the initial condition)</u> <u>such that</u> (A4) <u>holds for</u> $t \geq T_\varepsilon$. <u>Let</u> $\sup_{t>0,\varepsilon>0} E|\xi^\varepsilon(t)| < \infty$, <u>where the</u> <u>expectations are with respect to the stationary measure</u> $\nu^\varepsilon(\cdot)$. <u>Then</u> $\{\mu^\varepsilon(\cdot)\}$ <u>converges weakly to</u> $\mu(\cdot)$ <u>as</u> $\varepsilon \to 0$.

<u>Remarks</u>. In the proof of the last assertion of the theorem, we only need prove that $\{\mu^\varepsilon(\cdot)\}$ is tight, and the given tightness condition is used for this. In applications, it is often possible to prove a result such as $\overline{\lim_{t\to\infty}} E_x|x^\varepsilon(t)| \leq K$, where x is a bounded random variable and K does not depend on x for small ε. This implies the first sentence of the replacement for (A4) in the theorem. See also Theorem 6 in Section III.

<u>A criterion for</u> (A4). The tightness (A4) can be proved if certain stability assumptions (of the 'recurrence' type) are made on the limit $x(\cdot)$; in particular, if there is a Liapunov function $V(\cdot)$ of the sort frequently used to prove recurrence for $x(\cdot)$. An 'averaged' form of $V(\cdot)$ is used; the averaged function is obtained similarly to the way the perturbation $f^\varepsilon(\cdot)$ to $f(\cdot)$ is obtained in applications of Theorems 1 to 3. Owing to the mixing condition (B1) and to the smoothness conditions (B2), (B4), below, Theorem 5 is too strong for many applications. The conditions are weakened in Theorem 6 (Section III). Further comments and developments appear in [15].

Let us specialize to the form

$$\dot{x}^\varepsilon = F(x^\varepsilon,\xi^\varepsilon)/\varepsilon + G(x^\varepsilon,\xi^\varepsilon) + \overline{G}(x^\varepsilon) \quad , \tag{2.5}$$

where $\xi^\varepsilon(t) = \xi(t/\varepsilon^2)$ for some stationary process $\xi(\cdot)$. Let $EF(x,\xi) = EG(x,\xi) = 0$. This noise scaling is a common way of getting a wide band noise process, with 'bandwidth' parameter ε. Under various

sets of conditions ([5],[7],[14]) the differential generator of $x(\cdot)$ is given by

$$Af(x) = f_x'(x)\bar{G}(x) + \int_0^\infty E\left[f_x'(x)F(x,\xi(t))\right]_x' F(x,\xi(0))dt \quad . \tag{2.6a}$$

If $\xi(\cdot)$ is Markov and $F(\cdot,\cdot)$ not smooth, then we use the representation

$$Af(x) = f_x'(x)\bar{G}(x) + \int_0^\infty E\left[E_{\xi(0)} f_x'(x)F(x,\xi(t))\right]_x' F(x,\xi(0))dt \, , \tag{2.6b}$$

if the derivatives exist for $t > 0$ and are integrable.

We will require (B1)-(B4) in Theorem 5. The constant K might vary from usage to usage. These conditions as well as the extensions to be discussed in Theorem 6 below fit many common applications.

B1. $\xi(\cdot)$ <u>is a bounded, right continuous, stationary</u> ϕ-<u>mixing process</u> [16] <u>with</u> $\int_0^\infty \phi^{1/2}(t)dt < \infty$.

B2. $F(\cdot,\cdot)$, $G(\cdot,\cdot)$ <u>and</u> $\bar{G}(\cdot)$ <u>are continuous</u>, R^r <u>valued functions whose growth</u> (as $|x| \to \infty$) <u>is</u> $O(|x|)$. <u>The partial derivatives of</u> $F(\cdot,\xi)$ <u>up to order</u> 2 (<u>and of</u> $G(\cdot,\xi)$ <u>up to order</u> 1) <u>are bounded uniformly in</u> x,ξ, <u>and</u> $EF(x,\xi) \equiv 0 \equiv EG(x,\xi)$.

B3. <u>There is a diffusion process</u> $x(\cdot)$ <u>with differential generator</u> A <u>defined by</u> (2.6), <u>and which satisfies</u> (A1)-(A2). <u>There is a continuous Liapunov function</u> $0 \le V(x) \to \infty$ <u>as</u> $|x| \to \infty$, <u>and a</u> λ_0 <u>and</u> $\alpha_0 > 0$ <u>such that</u> $AV(x) \le -\alpha_0$ <u>for</u> $x \notin Q_0 = \{x: V(x) \le \lambda_0\}$. <u>The partial derivatives of</u> $V(\cdot)$ <u>up to order</u> 3 <u>are continuous</u>.

B4. <u>There are constants</u> K <u>such that</u>, <u>uniformly in</u> x,ξ,

$$|V_x'(x)G(x,\xi)| + |V_x'(x)F(x,\xi)| \le \big(1 + V(x)\big)$$

$$\big|\big(V_x'(x)F(x,\xi)\big)_x' \; F(x,\xi)\big| \le K\big(1 + V(x)\big)$$

$$\big|\big(V_x'(x)G(x,\xi)\big)_x' \; U(x,\xi)\big| \le K\big(1 + |AV(x)|\big)$$
$$\text{for} \quad U = F, G, \overline{G} \ .$$

$$\big|\big(V_x'(x)F(x,\xi)\big)_x' \; U(x,\xi)\big| \le K\big(1 + |AV(x)|\big), \quad U = G, \overline{G} \ .$$

$$\Big|\Big(\big(V_x'(x)F(x,\xi)\big)_x' \; F(x,\xi)\Big)_x' \; U(x,\xi)\Big| \le K\big(1 + |AV(x)|\big) \ ,$$
$$U = F, G, \overline{G} \ .$$

Theorem 5. Under (B1) - (B4) and the tightness of $\{x^\varepsilon(0)\}$, condition (A4) holds.

III. EXAMPLE 1. A PHASE LOCKED LOOP WITH A LIMITER

The phase locked loop (PLL) is a basic device used for synchronization in communications systems; in particular, to estimate the phase, frequency and certain 'timing' data of the incoming signal. In this example, we treat a simple PLL, but with a commonly used (and hard to analyze) nonlinearity. The circuit and definitions of some terms are given by Figure 1. The VCO (voltage controlled oscillator) is an oscillator which oscillates at a frequency whose deviation from a given central frequency is proportional to its input voltage. First, we describe the input noise and the parametrization, and then simplify the model slightly. Then Theorems 1 to 4 will be applied. The noise model used is chosen for ease of presentation, and because the technical details of proof [13] are not too hard. Further details of this example can be found in [13], which also contains analyses of several other phase locked loop type devices, with either nonlinearities or other features which seem to require a robust diffusion approximation method, such as the one presented here.

Refer to Figure 1. The object is to track the unknown and time varying phase $\theta(\cdot)$ of the input signal. Let $\theta(\cdot)$ be continuously differentiable here. With not much extra trouble, various random process models can be used for both $\theta(\cdot)$ and also for A_0 (which is a constant here). The system operates with a high central frequency and a noise bandwidth which is wide, but small relative to the central frequency. Let $\{q_\varepsilon\}$ be a sequence of positive real numbers tending to zero such that $\varepsilon/q_\varepsilon \to 0$ as $\varepsilon \to 0$. Let $\{\phi_i, z_i(\cdot), i=1,2\}$ be mutually independent with the ϕ_i uniformly distributed on $[0,2\pi]$, and the $z_i(\cdot)$ Gaussian processes with spectral density $S(\omega)$, and unit variance. Define $z_i^\varepsilon(t) = z_i(t/q_\varepsilon^2)$ and

$$\xi^\varepsilon(t) = \sigma[z_1^\varepsilon(t)\cos(\omega_0^\varepsilon t + \phi_1) + z_2^\varepsilon(t)\sin(\omega_0^\varepsilon t + \phi_2)]/q_\varepsilon , \qquad (3.1)$$

where ω_0^ε is the input (and central VCO) frequency. Then $\frac{1}{2}[S(q_\varepsilon^2(\omega-\omega_0^\varepsilon)) + S(q_\varepsilon^2(\omega+\omega_0))] = S_\varepsilon(\cdot)$ is the spectral density of $\xi^\varepsilon(\cdot)$. Let $\omega_0^\varepsilon = \omega_0/\varepsilon^2$. Then $S_\varepsilon(\cdot)$ is of the desired parametrized wide band form. With no loss of generality, we set $\phi_i = 0$. The matrix D in the filter in Fig. 1 is stable.

A gain of roughly the form $L_\varepsilon = L/q_\varepsilon$ in Fig. 1 (either before or after the filter) is needed because if $q_\varepsilon L_\varepsilon \to 0$ as $\varepsilon \to 0$, then the input to the VCO will go to zero weakly as $\varepsilon \to 0$. If $q_\varepsilon L_\varepsilon \to \infty$ as $\varepsilon \to 0$, then the output will diverge to infinity. The scaling is consistent with heuristic methods used for analyzing such systems. Let sign (0) be any number in $[-1,1]$. Let $z_i(\cdot)$ be Gauss-Markov; in particular let $z_i(t) = MZ_i(t)$, where $dZ_i = AZ_i dt + Bdw_i$, and $\{w_i(\cdot), i = 1,2\}$ are independent standard Wiener processes and A a stable matrix.

We now simplify the model slightly. The output of the multiplier can be written in the form

$$\frac{\sigma}{2q_\varepsilon}[z_1^\varepsilon(t)\left(\cos\hat\theta^\varepsilon + \cos(2\omega_0^\varepsilon t + \hat\theta^\varepsilon)\right) + z_2^\varepsilon(t)\left(\sin(-\hat\theta^\varepsilon) + \sin(2\omega_0^\varepsilon t + \hat\theta^\varepsilon)\right)]$$

$$+ \frac{A_0}{2}[\sin(\theta^\varepsilon - \hat\theta^\varepsilon) + \sin(\theta^\varepsilon + \hat\theta^\varepsilon + 2\omega_0^\varepsilon t)] \quad . \qquad (3.2)$$

If we proceeded with this 'output,' the terms involving $2\omega_0^\varepsilon t$ would have no effect on the limit, and for notational simplicity here we drop them at this point. In fact, it is common engineering practice to assume that the multiplier acts as a filter which eliminates these same terms involving $2\omega_0^\varepsilon t$. While one must be careful in using this 'filtering' idea, since the system is nonlinear and time varying and the properties of $\hat\theta^\varepsilon$ are not known *a priori* for small ε, these terms do

not, in fact, affect the limit process (as can be shown by retaining them throughout the analysis). Let $L = 1$ (absorb it into the other parameters).

The main result is the following: $\{v^{\varepsilon}(\cdot), \hat{\theta}^{\varepsilon}(\cdot)\}$ converges weakly to the diffusion $v(\cdot), \hat{\theta}(\cdot)$ defined by (3.3), where $B(\cdot)$ is a standard Wiener process and $\rho(t) = Ez_i(t)z_i(0)/Ez_i^2(t)$ (the stationary measure is to be used here). If $\rho(t) = \exp -a|t|$, $a > 0$, then $\sigma_0^2 = 2\ell n2/a$ [11], Section 6.

$$dv = [Dv + \frac{A_0}{\sigma} H \sqrt{2/\pi} \sin(\theta - \hat{\theta})]dt + H\sigma_0 dB \tag{3.3a}$$

$$d\hat{\theta} = Cvdt, \quad \dot{\theta} \text{ given,} \tag{3.3b}$$

$$\begin{aligned}\sigma_0^2 &= 4\int_0^{\infty} [P\{z_i(u) > 0, \; z_i(0) > 0\} - P\{z_i(u) < 0, \; z_i(0) > 0\}]du \\ &= \frac{4}{\pi}\int_0^{\infty} \sin^{-1}\rho(u)du\end{aligned}$$

If we consider the standard PLL, where the limiter and gain L_{ε} are replaced by simply a unity gain, then the limit process is defined by

$$\begin{aligned} dv &= [Dv + H\frac{A_0}{2}\sin(\theta - \hat{\theta})]dt + \sigma_1 HdB, \\ d\hat{\theta} &= Cvdt, \end{aligned} \tag{3.4}$$

where

$$\sigma_1^2 = \frac{\sigma^2}{2}\int_0^{\infty} \rho(u)du .$$

In this latter case, the analysis is much simpler and the Gaussian assumption can be weakened.

For small σ, the system with the limiter is preferable to the system without the limiter, and conversely for large σ. This interesting and very useful result has been verified also by simulations.

In order to use the limit results (3.3) or (3.4) to get probability estimates in any particular application, we must estimate the values of ω_0, σ^2, $\rho(\cdot)$ for that application. This can often be done via measurements on the system, and is briefly discussed in [14].

Application of Theorems 1 and 2. Dropping the terms in (3.2) which contain $2\omega_0^\varepsilon t$, the input to the filter in Fig. 1 is $u^\varepsilon\big(t,\hat{\theta}^\varepsilon(t),\theta(t)\big)$, where

$$u^\varepsilon(t,\hat{\theta},\theta) = \frac{1}{q_\varepsilon}\,\mathrm{sign}\Big[\frac{\sigma}{2q_\varepsilon}\Big(z_1^\varepsilon(t)\cos\hat{\theta} - z_2^\varepsilon(t)\sin\hat{\theta}\Big) + \frac{A_0}{2}\sin(\theta-\hat{\theta})\Big]\ .$$

Also

$$\begin{aligned} \dot{v}^\varepsilon &= Dv^\varepsilon + HEu^\varepsilon + H(u^\varepsilon - Eu^\varepsilon) \\ \dot{\hat{\theta}}^\varepsilon &= Cv^\varepsilon\ ,\quad (v^\varepsilon,\hat{\theta}^\varepsilon) = x^\varepsilon\ , \end{aligned} \tag{3.5}$$

$$Eu^\varepsilon(t,\hat{\theta},\theta) = \sqrt{2/\pi}\ \frac{A_0}{\sigma}\sin(\theta-\hat{\theta}) + O(q_\varepsilon)\ .$$

Throughout expectations and conditional expectations are over the $z_i^\varepsilon(\cdot)$ only. The $\hat{\theta}^\varepsilon(t)$, $\theta(t)$ are considered to be parameters when taking expectations.

Now, apply Theorem 1 to a test function $f(\cdot) \in \hat{\mathcal{C}}_0^{2,3}$. To be consistent with Theorems 1 and 2, we should work with $\{x^{\varepsilon,N}(\cdot)\}$, for large N. We will give only a rough outline of the procedure and the notation is simpler if we drop the N-affix and the $b_N(\cdot)$ (unless mentioned otherwise). The given estimates are actually valid for $\{x^{\varepsilon,N}(\cdot)\}$ for each N, and not necessarily for $\{x^\varepsilon(\cdot)\}$, but this is unimportant for application of Theorems 1 and 2. Now, applying $\hat{A}^\varepsilon$ to $f(\cdot)$, and writing $x^\varepsilon(t) = x^\varepsilon$, yields

$$\hat{A}^{\varepsilon}f(x^{\varepsilon},t) = f_t(x^{\varepsilon},t) + f_{\hat{\theta}}(x^{\varepsilon},t)Cv^{\varepsilon} + f_v'(x^{\varepsilon},t)[Dv^{\varepsilon} + HEu^{\varepsilon}(t,\hat{\theta}^{\varepsilon},\theta(t))] + f_v'(x^{\varepsilon},t)H\delta u^{\varepsilon}(t,\hat{\theta}^{\varepsilon},\theta(t)) \quad ,$$

where $\delta u^{\varepsilon}(t,\hat{\theta},\theta) = u^{\varepsilon}(t,\hat{\theta},\theta) - Eu^{\varepsilon}(t,\hat{\theta},\theta)$. Only the 'noise term' $f_v'H\delta u^{\varepsilon}$ needs to be averaged out. These so-called averaging operations follow the scheme in [3], [7] and they yield the perturbations f_i^{ε} to f required by Theorems 1, 2. (See also [9], [12]-[15], for applications and illustrations of this averaging idea.)

Define the first perturbation $f_1^{\varepsilon}(t) = f_1^{\varepsilon}(x^{\varepsilon}(t),t)$ by

$$\begin{aligned} f_1^{\varepsilon}(x^{\varepsilon},t) &= \int_0^{\infty} dsE_t^{\varepsilon}f_v'(x^{\varepsilon},t+s)H\delta u^{\varepsilon}(t+s,\hat{\theta}^{\varepsilon},\theta(t)) \\ &= q_{\varepsilon}^2\int_0^{\infty} dsE_t^{\varepsilon}f_v'(x^{\varepsilon},t+q_{\varepsilon}^2s)H\delta u^{\varepsilon}(t+q_{\varepsilon}^2s,\hat{\theta}^{\varepsilon},\theta(t))ds \end{aligned} \tag{3.6}$$

Note that the integrand of the first integral of (3.6) at $s=0$ is just the 'noise term' which we wish to average out. Owing to the sign function in δu^{ε}, and to the fact that the operation $\hat{A}^{\varepsilon}f_1^{\varepsilon}$ involves a differentiation, it is convenient to write $E_t^{\varepsilon}\delta u^{\varepsilon}$ as follows. Define $[t+s] \equiv [z_1^{\varepsilon}(t+s)\cos\hat{\theta}^{\varepsilon} - z_2^{\varepsilon}(t+s)\sin\hat{\theta}^{\varepsilon} + \frac{q_{\varepsilon}A_0}{2}\sin(\theta-\hat{\theta}^{\varepsilon})]$. The distribution of $[t+s]$ conditioned on $\{z_i^{\varepsilon}(u),\ u\leq t,\ i=1,2\}$ is $N(m_{\varepsilon}(s),\ \sigma_{\varepsilon}^2(s))$, where

$$\begin{aligned} m_{\varepsilon}(s) &= \rho(s/q_{\varepsilon}^2)[z_1^{\varepsilon}(t)\cos\hat{\theta}^{\varepsilon} - z_2^{\varepsilon}(t)\sin\hat{\theta}^{\varepsilon}] + \frac{A_0q_{\varepsilon}}{2}\sin(\theta-\hat{\theta}^{\varepsilon}) \\ \sigma_{\varepsilon}^2(s) &= 1 - \rho^2(s/q_{\varepsilon}^2) \quad . \end{aligned}$$

Then, using the notation $dN(m,\sigma^2) = d\xi[\exp-(\xi-m)^2/2\sigma^2]/\sqrt{2\pi}\,\sigma$, we have

$$E_t^{\varepsilon}\delta u^{\varepsilon}(t+s,\hat{\theta}^{\varepsilon},\theta(t))$$

$$= \frac{1}{q_{\varepsilon}}\int_{-\infty}^{\infty}(\text{sign }\xi)[dN(m_{\varepsilon}(s),\sigma_{\varepsilon}^2(s)) - dN(\frac{A_0q_{\varepsilon}}{\sigma}\sin(\theta(t)-\hat{\theta}^{\varepsilon}),1)] \quad . \tag{3.7}$$

It can be shown that $f_1^\varepsilon(\cdot) \in \mathcal{D}(\hat{A}^\varepsilon)$, and that

$$|f_1^\varepsilon(t)| = O(q_\varepsilon)[1 + |z^\varepsilon(t)|] \quad , \tag{3.8}$$

and (writing $x^\varepsilon(t) = x^\varepsilon$, and letting I denote the integrand of the first integral in (3.6))

$$\hat{A}^\varepsilon f_1^\varepsilon(x^\varepsilon,t) = -f_v'(x^\varepsilon,t)H\delta u^\varepsilon\big(t,\hat{\theta}^\varepsilon,\theta(t)\big) + \int_0^\infty ds(E_t^\varepsilon I)_{\hat{\theta}}\dot{\hat{\theta}}^\varepsilon +$$
$$+ \int_0^\infty ds(E_t^\varepsilon I)_\theta \dot{\theta}(t) + \int_0^\infty ds(E_t^\varepsilon I)_v' \dot{v}^\varepsilon \quad . \tag{3.9}$$

In (3.9), the representation (3.7) is used for $E_t^\varepsilon \delta u^\varepsilon$, so all derivatives are taken with respect to parameters in the normal density function. The first two integrals on the right of (3.9) are bounded by the right side of (3.8) (to see this, change variables $s/q_\varepsilon^2 \to s$), and the last term equals (3.10) modulo a term which is bounded by the right side of (3.8). In (3.10), $[t+s]$ denotes the term defined below (3.6).

$$\frac{1}{q_\varepsilon^2}\int_0^\infty dsH'f_{vv}(x^\varepsilon,t+s)H(E_t^\varepsilon \text{sign}[t+s] - E\,\text{sign}\,[t+s])(\text{sign}\,[t] - E\,\text{sign}\,[t])$$
$$\equiv k^\varepsilon(t,x^\varepsilon) \tag{3.10}$$

The term k^ε is neither negligible nor part of $Af(x^\varepsilon,t)$ and must be averaged out. To do this, we define the second perturbation function $f_2^\varepsilon(t) = f_2^\varepsilon\big(t,x^\varepsilon(t)\big)$, where

$$f_2^\varepsilon(t,x^\varepsilon) = \int_0^\infty d\tau[E_t^\varepsilon k^\varepsilon(t+\tau,x^\varepsilon) - Ek^\varepsilon(t+\tau,x^\varepsilon)] \quad .$$

It can be shown that $|f_2^\varepsilon(t)| = O(q_\varepsilon)[1 + |z^\varepsilon(t)|^2]$, that $f_2^\varepsilon(\cdot) \in \mathcal{D}(\hat{A}^\varepsilon)$ and that (again writing $x^\varepsilon = x^\varepsilon(t)$)

$$\hat{A}^\varepsilon f_2^\varepsilon(t,x^\varepsilon) = -k^\varepsilon(t,x^\varepsilon) + Ek^\varepsilon(t,x^\varepsilon) + \text{error term} \;, \tag{3.11}$$

where $|\text{error term}| = O(q_\varepsilon)[1+|z^\varepsilon(t)|^3]$. In applying $\hat{A}^\varepsilon$ to $f_2^\varepsilon(\cdot)$ we use a representation of the type used in (3.7) for the conditional expectations and expectations, so that the derivatives are taken only with respect to parameters in the normal density functions. Finally, it can be shown via the technique in [11, Section 6] that $Ek^\varepsilon(t,x) - H'f_{vv}(x,t)H\sigma_0^2/2 \to 0$ uniformly in bounded (x,t) sets as $\varepsilon \to 0$.

Now, let us summarize the above calculations, and reintroduce the superscript N on $x^{\varepsilon,N}(\cdot)$ only. The stated bounds on $|f_i^\varepsilon(\cdot)|$, $|f^\varepsilon(\cdot)-f(\cdot)|$, and on the error term in (3.11) and on the first two integrals on the right side of (3.9) all hold for each N. Now Theorem 2 implies tightness for $\{x^{\varepsilon,N}(\cdot)\}$ for each N, and $\underset{\varepsilon\to 0}{p\text{-}\lim}[f^\varepsilon(\cdot)-f(x^{\varepsilon,N}(\cdot),\cdot)] = 0$. The expressions given for $\hat{A}^\varepsilon f_i^\varepsilon(x^\varepsilon,t)$ above equal $\hat{A}^\varepsilon f_i^\varepsilon(x^{\varepsilon,N},t)$, if $x^{\varepsilon,N}$ and x^ε are in S_N. Thus,

$$\underset{\varepsilon\to 0}{p\text{-}\lim}[\hat{A}^\varepsilon f^\varepsilon(x^{\varepsilon,N}(\cdot),\cdot) - \left(\frac{\partial}{\partial t} + A\right)f(x^{\varepsilon,N}(\cdot),\cdot)]b_N(x^{\varepsilon,N}(\cdot)) = 0 , \quad (3.12)$$

where A is the differential generator of the process (3.3), and we conclude that (3.3) is the weak limit of $\{v^\varepsilon(\cdot),\hat{\theta}^\varepsilon(\cdot)\}$ if $\{v^\varepsilon(0),\hat{\theta}^\varepsilon(0)\}$ converge weakly to $\{v(0),\hat{\theta}(0)\}$. Further details on this and on similar problems are in [11], [12], and [14].

<u>Approximate invariant measures for</u> (3.5). Since the noise $z^\varepsilon(\cdot)$ is unbounded, Theorem 5 cannot be applied, but there is a result similar to Theorem 4 which is useful. Let $\theta(t) = \theta_0$, a constant. We interpret both $\hat{\theta}^\varepsilon(\cdot)$ and $\hat{\theta}(\cdot)$ modulo 2π, so they are always bounded. Also conditions (A1)-(A3) hold. The main problem is in showing that $\{v^\varepsilon(\cdot),\ t \geq 0,\ \varepsilon \text{ small}\}$ is tight. In particular, we want to show that $(\hat{\theta}^\varepsilon(\cdot),v^\varepsilon(\cdot)) = x^\varepsilon(\cdot)$ has an invariant measure $m^\varepsilon(\cdot)$ for small ε, and that the v-marginals $\{\mu^\varepsilon(\cdot)\}$ of $\{m^\varepsilon(\cdot)\}$ are tight.

There is a useful 'averaged Liapunov method' which is similar to Theorem 5, but where the noise is allowed to be unbounded, and the conditions are tailored to a class of cases which occurs frequently. For motivation, we first consider the present case. For some positive definite and symmetric P, let $V(x) = v'Pv$ be a Liapunov function for $\dot{v} = Dv$, and let $\mathcal{X}$ denote the differential generator of (3.3a). Then there are $\gamma > 0$, $K < \infty$ such that $\mathcal{X}V(x) \leq -\gamma V(x) + K$. An averaging method can now be applied to $V(\cdot)$, to yield a perturbed Liapunov function $V^{\varepsilon}(\cdot)$, which in turn is useful for proving the required tightness, via a method which resembles those used for 'Liapunov function' proofs of recurrence of strong Markov processes. Actually, since $\hat{A}^{\varepsilon}$ cannot be applied to unbounded functions, the averaging is done on and the technical conditions are introduced on the truncated Liapunov function defined by $V_M(x) = V(x)b_M(v)$, where $b_M(\cdot)$ is the truncation function defined earlier.

The perturbed or averaged truncated Liapunov function $V_M^{\varepsilon}(t) = V_M^{\varepsilon}(x^{\varepsilon}(t),t)$ is obtained by perturbing $V_M(v^{\varepsilon}(t))$, in exactly the same way that we perturbed $f(x^{\varepsilon,N}(t),t)$ to get $f^{\varepsilon}(x^{\varepsilon,N}(t),t) = f^{\varepsilon}(t)$, except that here, we use the original process $x^{\varepsilon}(\cdot)$ and not the truncation $x^{\varepsilon,N}(\cdot)$.

In more generality, the technique, assumptions and result are formalized as follows. We suppose that $\dot{x}^{\varepsilon} = H^{\varepsilon}(x^{\varepsilon},\xi^{\varepsilon})$ takes the form (2.5) and that $\xi^{\varepsilon}(t) = \xi(t/\varepsilon^2)$. (In the example of this section, q_{ε} is used in lieu of ε.) Suppose that for some $\varepsilon_0 > 0$, $V_M^{\varepsilon}(\cdot) \in \mathcal{D}(\hat{A}^{\varepsilon})$ for $\varepsilon \leq \varepsilon_0$. Suppose that there are random variables $\tilde{\xi}^{\varepsilon}(t),\bar{\xi}^{\varepsilon}(t)$, integers p,q, and functions $\tilde{V}(\cdot),\overline{V}(\cdot)$ satisfying (3.13). For the first two lines of (3.13), let $x \in S_M = \{x : |x| \leq M\}$. Let A be defined by (2.6).

$$\hat{A}^{\varepsilon}V_M^{\varepsilon}(x,t) = AV(x) + O(\varepsilon)\big(1+V(x)\big) + O(\varepsilon)\tilde{\xi}^{\varepsilon}(t)\tilde{V}(x,t)$$

$$V_M^{\varepsilon}(x,t) = V_M(x) + O(\varepsilon)\big(1+V(x)\big) + O(\varepsilon)\bar{\xi}^{\varepsilon}(t)\bar{V}(x,t) \qquad (3.13)$$

$$\sup_{t,\varepsilon} E|\tilde{\xi}^{\varepsilon}(t)|^p < \infty \quad , \quad \sup_{t,\varepsilon} E|\bar{\xi}^{\varepsilon}(t)|^q < \infty$$

$$|\tilde{V}(x,t)|^{p/p-1} = O\big(V(x)\big) \quad , \quad |\bar{V}(x,t)|^{q/q-1} = O\big(V(x)\big) \quad , \quad \text{for large } |x| .$$

We also need (C1)-(C4).

C1. $\bar{G}(\cdot)$, $F(\cdot,\cdot)$, $G(\cdot,\cdot)$ are measurable and are $O(|x|)$ for large $|x|$, uniformly in bounded ξ-sets.

C2. For the given sequence $\{x^{\varepsilon}(0)\}$, $\sup_{\varepsilon\le\varepsilon_0} EV\big(x^{\varepsilon}(0)\big) < \infty$.

C3. $\sup_t E|\xi(t)| < \infty$. For small ε, $\{x^{\varepsilon}(\cdot),\xi^{\varepsilon}(\cdot)\}$ and $\xi(\cdot)$ are Markov-Feller processes with right continuous paths and homogeneous transition functions.

C4. $0 \le V(x) \to \infty$ as $|x| \to \infty$ and $AV(x) \le -\gamma V(x)$ for some $\gamma > 0$ and large $|x|$.

Theorem 6 [15]. Assume (C1)-(C4), (A3), (B3), (3.13) and the conditions above (3.13). Then there is an $\varepsilon_1 > 0$ such that for $\varepsilon \le \varepsilon_1$, $\big(x^{\varepsilon}(\cdot),\ \xi^{\varepsilon}(\cdot)\big)$ has an invariant measure $m^{\varepsilon}(\cdot)$. The x-marginals $\{\mu^{\varepsilon}(\cdot)\}$ of any such sequence of invariant measures converge weakly to $\mu(\cdot)$ as $\varepsilon \to 0$.

Application of Theorem 6 to Example 1. Let $\theta(t) \equiv \theta(0)$. Since $\hat{\theta}(\cdot)$ and $\hat{\theta}^{\varepsilon}(\cdot)$ are bounded by 2π, we need only let $V(\cdot)$ depend on v. Use $V(x) = v'Pv$, and q_{ε} in lieu of ε. Then we can get the form (3.13) where $p = q = 2$, $\overline{V}(x,t) = O(|v| + 1)$, $\tilde{V}(x,t) = O(|v| + 1)$ and both $\tilde{\xi}^{\varepsilon}(t)$ and $\bar{\xi}^{\varepsilon}(t)$ are bounded by $O(1 + |z^{\varepsilon}(t)|^3)$. Then the theorem implies that $\{x^{\varepsilon}(\cdot), \xi^{\varepsilon}(\cdot)\}$ has an invariant measure for small $\varepsilon > 0$, and that the x-marginals converge to $\mu(\cdot)$, the unique invariant measure of $x(\cdot) = (v(\cdot), \hat{\theta}(\cdot))$, as $\varepsilon \to 0$. This justifies the use of the stationary measure of (3.3) as an approximation to the stationary measures of the system of Fig. 1, for small enough ε, and when $\theta(t) \equiv \theta_0$.

IV. EXAMPLE 2. A DIGITAL PHASE LOCKED LOOP (DPLL)

Consider the digital signalling problem where a random sequence $\{s_n\}$ is transmitted. The received signal is $s(t) + \text{noise}(t) \equiv y_T(t)$, where $s(t) = s_n$ in $[nT + \delta_0, nT + T + \delta_0)$. The time T is the symbol interval and δ_0 is unknown. For detection with an acceptably low error rate, a very good estimate of δ_0 is needed, and a DPLL system is used to provide a sequence of estimates of δ_0. See, for example, the systems in [17]. Here we deal with a specific simple (but important) case. Let $s_n = \pm A_0$, $A_0 > 0$, where $\{s_n\}$ is i.i.d. and $P\{s_n = A_0\} = 1/2$. Let $\hat{\delta}_n$ denote the n*th* estimate of δ_0, define $\lambda_n = (\hat{\delta}_n - \delta_0)/T$. In the study of the algorithm, the exact value of δ_0 is unimportant, so w.l.o.g., set $\delta_0 = 0$.

The particular DPLL studied will now be defined. It was chosen because variations of it are used frequently [17], and rigorous analysis via currently used techniques in communication theory is very hard. The diffusion approximation method used here involves scalings which are consistent with those used currently in communication theory (explicitly or implicitly), and seems to be a robust and systematic tool for this type of problem. Let $\text{noise}(\cdot) = \psi_T(\cdot)$, where $w_T(t) \equiv \int_0^t \psi_T(s)ds$ is a Wiener process with variance $\sigma_T^2 t$. The variance σ_T^2 must depend on T (for small T) if the problem is to be meaningful. In any particular case, of course, σ_T^2 would have to be determined from the problem data. To get the scaling of σ_T^2, suppose that $\hat{\delta}_n = 0$ and that the statistic $y = s_0 T + \int_0^T \psi_T(s)ds$ (the integral of the observation, over an interval of the same width as the symbol interval) is used to estimate $s_0 (= \pm A_0)$ via a likelihood ratio test. But,

$$P\{s_0 = A_0 \text{ chosen} \mid s_0 = -A_0\} = \int_{A_0T/\sigma_T\sqrt{T}}^{\infty} dN(0,1) \quad .$$

Thus $\sigma_T^2 = \sigma^2 T$ for some $\sigma^2 > 0$ is the natural scaling if the decision problem is not to degenerate for small $T > 0$.

The algorithm (DPLL) is defined as follows. Let $0 < \Delta < 1/4$. Define $e_n(\lambda_{n-1},\lambda_n)$ by

$$e_n(\lambda_{n-1},\lambda_n) = \left|\int_{(n+\Delta)T+\hat{\delta}_{n-1}}^{(n+1+\Delta)T+\hat{\delta}_n} y_T(s)ds\right| - \left|\int_{(n+1-\Delta)T+\hat{\delta}_{n-1}}^{(n+2-\Delta)T+\hat{\delta}_n} y_T(s)ds\right| \quad . \tag{4.1}$$

Thus we can write, for a Wiener process $W(\cdot)$ with variance $\sigma^2 t$,

$$g_n(\lambda_{n-1},\lambda_n) \equiv e_n(\lambda_{n-1},\lambda_n)/T = |(1-\Delta-\lambda_{n-1})s_n + (\Delta+\lambda_n)s_{n+1} + W(n+1+\Delta+\lambda_n) - W(n+\Delta+\lambda_{n-1})|$$

$$-|(\Delta-\lambda_{n-1})s_n + (1-\Delta+\lambda_n)s_{n+1} + W(n+2-\Delta+\lambda_n) - W(n+1-\Delta+\lambda_{n-1})| \quad . \tag{4.2}$$

We can use $W(t)$ instead of $w_T(tT)/T$, since the latter process is a Wiener process with variance $\sigma^2 t$. Finally, with adjustment parameter $\varepsilon > 0$, we write λ_n as λ_n^ε and let $\{\lambda_n^\varepsilon\}$ be defined by the recursive algorithm (the DPLL):

$$\lambda_{n+1}^\varepsilon = \lambda_n^\varepsilon + \varepsilon g_n(\lambda_{n-1}^\varepsilon,\lambda_n^\varepsilon) \quad , \quad \varepsilon > 0 \quad . \tag{4.3}$$

Normally, the adjustment parameter ε depends on T. Since we are concerned with tracking over real time intervals $\{n : nT \le t\}$, all real t, ε should be $\approx$ (constant)T for small T for otherwise (smaller order of T), the sequence $\{\lambda_n\}$ moves too slowly, or (larger order of T) is excessively sensitive to the noise. Let $\lambda_0^\varepsilon = O(\sqrt{\varepsilon})$, for otherwise, the system (4.3) will not be able to improve the estimate for small ε.

Define $U_n^\varepsilon = \lambda_n^\varepsilon/\sqrt{\varepsilon}$, and define $U^\varepsilon(t) = U_n^\varepsilon$ on $[n\varepsilon, n\varepsilon+\varepsilon)$. Using parameters λ,λ' (in lieu of λ_{n-1},λ_n), define $\bar{g}(\lambda,\lambda') = Eg_n(\lambda,\lambda')$, and

set $\xi_n^\varepsilon(\lambda_{n-1}^\varepsilon,\lambda_n^\varepsilon) = g_n(\lambda_{n-1}^\varepsilon,\lambda_n^\varepsilon) - \bar{g}(\lambda_{n-1}^\varepsilon,\lambda_n^\varepsilon)$. Then

$$\begin{aligned} U_{n+1}^\varepsilon &= U_n^\varepsilon + \varepsilon\left(\bar{g}(\lambda_{n-1}^\varepsilon,\lambda_n^\varepsilon)/\sqrt{\varepsilon}\right) + \sqrt{\varepsilon}\,\xi_n^\varepsilon(\lambda_{n-1}^\varepsilon,\lambda_n^\varepsilon) \\ &\equiv U_n^\varepsilon + \varepsilon q_\varepsilon(\lambda_{n-1}^\varepsilon,\lambda_n^\varepsilon) \quad . \end{aligned} \tag{4.4}$$

Define $U_n^{\varepsilon,N}, U^{\varepsilon,N}(\cdot)$ as $x^{\varepsilon,N}(\cdot)$ was defined in Section II. In particular, $U_{n+1}^{\varepsilon,N} = U_n^{\varepsilon,N} + \varepsilon q_\varepsilon(\lambda_{n-1}^{\varepsilon,N},\lambda_n^{\varepsilon,N})b_N(U_n^{\varepsilon,N})$, where $\lambda_n^{\varepsilon,N} = \sqrt{\varepsilon}\,U_n^{\varepsilon,N}$ is used. There is a $\theta > 0$ such that $-\theta = (d/d\lambda)\bar{g}(\lambda,\lambda)\big|_{\lambda=0}$. Equation (4.4) can be written as

$$U_{n+1}^\varepsilon = U_n^\varepsilon - \theta U_n^\varepsilon + \sqrt{\varepsilon}\,\xi_n^\varepsilon(\lambda_{n-1}^\varepsilon,\lambda_n^\varepsilon) + \varepsilon v_n^\varepsilon \quad , \tag{4.5}$$

where $v_n^\varepsilon = O(|\lambda_n^\varepsilon|^2+|\lambda_{n-1}^\varepsilon|^2+|U_n^\varepsilon-U_{n-1}^\varepsilon|)$, and similarly for the $\{U_n^{\varepsilon,N}\}$ equation. The main result is that if U_0^ε converges weakly to a random variable U_0, then $\{U^\varepsilon(\cdot)\}$ converges weakly to $U(\cdot)$, where

$$\begin{aligned} dU &= -\theta U dt + v dB, \quad U(0) = U_0 \quad , \\ v^2 &= 2E[g_{n+1}(0,0) - \bar{g}(0,0)][g_n(0,0) - \bar{g}(0,0)] \\ &\quad + E[g_n(0,0) - \bar{g}(0,0)]^2 \quad , \quad \text{any} \quad n \geq 1 \ . \end{aligned} \tag{4.6}$$

If the $\{v_n^\varepsilon\}$ were carried through in the calculations, they would contribute nothing to the limit, so for notational simplicity here, we drop them from (4.5) (and from the equation for $\{U_n^{\varepsilon,N}\}$) now.

The representation (4.6) is useful for calculating the statistics of the DPLL (4.3) when the adjustment rate ε is small, say $\varepsilon \simeq T$, where the symbol interval is small. In any case, rough qualitative properties and parametric dependencies can be obtained. We do not know how small ε must be for (4.6) to be a good (scaled) approximation,

although simulations on related systems suggest that the range is within reason for practical systems.

Application of Theorem 3. To get the limit (4.6), use Theorem 3. Fix $f(\cdot) \in \hat{\mathcal{C}}_0^{2,3}$. The method of getting and using the $f^{\varepsilon,N}(\cdot)$ (written here as $f^{\varepsilon}(\cdot)$, for notational simplicity) is very similar to that for continuous parameter case. We start by calculating $\hat{A}^{\varepsilon} f\big(U^{\varepsilon,N}(n\varepsilon)\big)$, and successively average out the 'noise terms,' except that the f_i^{ε} are defined by sums rather than integrals. More detail on the method for this and related DPLL's in [13], and details of similar approximations for other scaled and interpolated discrete parameter processes can be found in [9], [19].

First, we calculate $\hat{A}^{\varepsilon} f(U_n^{\varepsilon,N}, n\varepsilon)$ (discrete parameter form). Let $E_n^{\varepsilon}(\cdot)$ = expectation conditioned on $W(t)$, $s(t)$, $t \le$ time at which λ_n^{ε} is first known.

$$
\begin{aligned}
\varepsilon \hat{A}^{\varepsilon} f(U_n^{\varepsilon,N}, n\varepsilon) &= \varepsilon f_t(U_n^{\varepsilon,N}, n\varepsilon) + O(\varepsilon^2) - \varepsilon f_u(U_n^{\varepsilon,N}, n\varepsilon)\theta U_n^{\varepsilon,N} b_N(U_n^{\varepsilon,N}) \\
&+ \sqrt{\varepsilon}\, f_u(U_n^{\varepsilon,N}, n\varepsilon) b_N(U_n^{\varepsilon,N}) E_n^{\varepsilon} \xi_n^{\varepsilon}(\lambda_{n-1}^{\varepsilon,N}, \lambda_n^{\varepsilon,N}) \qquad (4.7) \\
&+ \varepsilon \frac{f_{uu}(U_n^{\varepsilon,N}, n\varepsilon)}{2} b_N^2(U_n^{\varepsilon,N}) E_n^{\varepsilon}\big(\xi_n^{\varepsilon}(\lambda_{n-1}^{\varepsilon,N}, \lambda_n^{\varepsilon,N})\big)^2 + O_{1n}^{\varepsilon,N} .
\end{aligned}
$$

The $O_{1n}^{\varepsilon,N}$, a remainder in the truncated Taylor expansion, satisfies

$$
O_{1n}^{\varepsilon,N} = O(\varepsilon^{3/2}) E_n^{\varepsilon}[1 + |\xi_n^{\varepsilon}(\lambda_{n-1}^{\varepsilon,N}, \lambda_n^{\varepsilon,N})|^3] . \qquad (4.8)
$$

Using the properties of the Wiener process, for each $t < \infty$,

$$
\begin{aligned}
&\lim_{\varepsilon \to 0} \sup_{n\varepsilon \le t} |O_{1n}^{\varepsilon,N}/\varepsilon| = 0 \quad \text{w.p.1}, \\
&\lim_{\varepsilon \to 0} \sup_{n\varepsilon \le t} E|O_{1n}^{\varepsilon,N}/\varepsilon| = 0 .
\end{aligned} \qquad (4.9)
$$

All the $\{O^{\varepsilon}_{in}\}$ introduced below satisfy (4.9) for each N.

Only the first and third terms on the right of (4.7) are part of the limit operator $\varepsilon(\partial/_{\partial t} + A)f(U^{\varepsilon,N}_n, n\varepsilon)$ (for $U^{\varepsilon,N}_n \in S_N$). The rest are either negligible or must be averaged out. Henceforth, drop the N superscript on $U^{\varepsilon,N}_n$, $\lambda^{\varepsilon,N}_n$. The bounds on the error terms are valid for each $\{U^{\varepsilon,N}_n, \varepsilon > 0\}$ as required by Theorem 3, but not necessarily for $\{U^{\varepsilon}_n, \varepsilon > 0\}$. We use the form $f^{\varepsilon}(n\varepsilon) = f(U^{\varepsilon}_n, n\varepsilon) + \sum_{i=0}^{2} f^{\varepsilon}_i(n\varepsilon)$, where it will be true that (each $t > 0$)

$$
\begin{aligned}
&\lim_{\varepsilon\to 0} \sup_{n\varepsilon\le t} |f^{\varepsilon}_i(n\varepsilon)| = 0 \quad \text{w.p.1},\\
&\lim_{\varepsilon\to 0} \sup_{n\varepsilon\le t} E|f^{\varepsilon}_i(n\varepsilon)| = 0 .
\end{aligned} \tag{4.10}
$$

To average out the f_{uu} term in (4.7), introduce the first perturbation $f^{\varepsilon}_0(n\varepsilon) = f^{\varepsilon}_0(U^{\varepsilon}_n, n\varepsilon)$, where

$$
f^{\varepsilon}_0(U,n\varepsilon) = \frac{\varepsilon\, f_{uu}(U,n\varepsilon)}{2} b^2_N(U) \sum_{j=n}^{\infty} [E^{\varepsilon}_n(\xi^{\varepsilon}_j(\lambda^{\varepsilon}_{n-1},\lambda^{\varepsilon}_n))^2 - E(\xi^{\varepsilon}_j(\lambda^{\varepsilon}_{n-1},\lambda^{\varepsilon}_n))^2] ,
$$

The $\lambda^{\varepsilon}_{n-1}, \lambda^{\varepsilon}_n$ are taken as parameters in calculating all expectations. Due to $\Delta \le 1/4$, and to the properties of the Wiener process $W(\cdot)$ and of $\{s_n\}$, all the terms in the sum are zero except the lowest (since $E^{\varepsilon}_n(\ ) = E(\ )$ for the other terms). Now

$$
\begin{aligned}
\varepsilon\hat{A}^{\varepsilon} f^{\varepsilon}_0(U^{\varepsilon}_n, n\varepsilon) &= -\,(5th \text{ term of } (4.7)) + \text{terms satisfying } (4.9)\\
&\quad + \frac{\varepsilon f_{uu}(U^{\varepsilon}_n, n\varepsilon)}{2} b^2_N(U^{\varepsilon}_n) E\xi^{\varepsilon}_n(\lambda^{\varepsilon}_{n-1},\lambda^{\varepsilon}_n)^2
\end{aligned}
$$

The last component yields the last term of the variance v^2 in (4.6).

To average out the $\sqrt{\varepsilon}$ term in (4.7) use

$$f_1^\varepsilon(n\varepsilon) = \sqrt{\varepsilon}\, f_u(U_n^\varepsilon, n\varepsilon) b_N(U_n^\varepsilon) \sum_{j=n}^{\infty} E_n^\varepsilon \xi_j^\varepsilon(\lambda_{n-1}^\varepsilon, \lambda_n^\varepsilon)$$
$$= \sqrt{\varepsilon}\, f_u(U_n^\varepsilon, n\varepsilon) b_N(U_n^\varepsilon) E_n^\varepsilon \xi_n^\varepsilon(\lambda_{n-1}^\varepsilon, \lambda_n^\varepsilon) \quad .$$

We can write

$$\varepsilon \hat{A}^\varepsilon f_1^\varepsilon(n\varepsilon) = -f_1^\varepsilon(n\varepsilon) + 0_{2n}^\varepsilon$$
$$+ \sqrt{\varepsilon}\, E_n^\varepsilon b_N(U_{n+1}^\varepsilon) f_u(U_{n+1}^\varepsilon, n\varepsilon) \xi_{n+1}^\varepsilon(\lambda_n, \lambda_{n+1}) \quad .$$

The last term can be written as

$$\sqrt{\varepsilon}\, E_n^\varepsilon \big(f_u(U_n^\varepsilon, n\varepsilon) b_N(U_n^\varepsilon)\big)_u (U_{n+1}^\varepsilon - U_n^\varepsilon) \xi_{n+1}^\varepsilon(\lambda_n^\varepsilon, \lambda_{n+1}^\varepsilon) + 0_{3n}^\varepsilon$$
$$= \varepsilon \big(f_u(U_n^\varepsilon, n\varepsilon) b_N(U_n^\varepsilon)\big)_u E_n^\varepsilon \xi_{n+1}^\varepsilon(\lambda_n^\varepsilon, \lambda_n^\varepsilon)\ \xi_n^\varepsilon(\lambda_{n-1}^\varepsilon, \lambda_n^\varepsilon) + 0_{4n}^\varepsilon \quad . \tag{4.11}$$

This term is neither negligible nor part of $\varepsilon A f(U_n^\varepsilon)$, and so it needs to be averaged further. This is done via the last perturbation $f_2^\varepsilon(n\varepsilon)$:

$$\varepsilon \big(f_u(U_n^\varepsilon, n\varepsilon) b_N(U_n^\varepsilon)\big)_u \sum_{j=n}^{\infty} [E_n^\varepsilon \xi_{j+1}^\varepsilon(\lambda_n^\varepsilon, \lambda_n^\varepsilon) \xi_j^\varepsilon(\lambda_{n-1}, \lambda_n)$$
$$- E\xi_{j+1}^\varepsilon(\lambda_n^\varepsilon, \lambda_n^\varepsilon) \xi_j^\varepsilon(\lambda_{n-1}^\varepsilon, \lambda_n^\varepsilon)] = f_2^\varepsilon(n^\varepsilon)$$

$\varepsilon \hat{A}^\varepsilon f_2^\varepsilon(n\varepsilon)$ can be shown to equal 0_{5n}^ε - (4.11) + $\frac{\varepsilon}{2}$ $\big(f_u(U_n^\varepsilon, n\varepsilon) b_N(U_n^\varepsilon)\big) \times$ (first component of variance v^2 in (4.6)$\big)$. The constructed $f^\varepsilon(n\varepsilon)$ can be shown to satisfy the requirements of Theorem 3, where the operator A is that of (4.6). Also, the $f^\varepsilon(n\varepsilon)$ can be used in Theorem 2 to yield tightness.

REFERENCES

1. E. Wong and M. Zakai, "On the convergence of ordinary integrals to stochastic integrals," *Ann. Math. Stat.*, 36, 1560-1564 (1965).
2. G. C. Papanicolaou and W. Kohler, "Asymptotic theory of mixing stochastic ordinary differential equations," *Comm. Pure and Appl. Math.*, 29, 641-668 (1974).
3. G. Blankenship and G. C. Papanicolaou, "Stability and control of stochastic systems with wide-band noise disturbances," *SIAM J. on Applied Math.*, 34, 437-476 (1978).
4. T. G. Kurtz, "Semigroups of conditional shifts and approximation of Markov processes," *Ann. Prob.*, 4, 618-642 (1975).
5. T. G. Kurtz, *Approximation of Population Processes* (Philadelphia: SIAM, 1981).
6. R. Z. Khazminskii, "A limit theorem for solutions of differential equations with random right-hand sides," *Theory Prob. and Applic.*, 11, 390-406 (1966).
7. H. J. Kushner, "Jump-diffusion approximations for ordinary differential equations with wide-band random right-hand sides," *SIAM J. on Control and Optimization*, 17, 729-744 (1979).
8. H. J. Kushner, "A martingale method for the convergence of a sequence of processes to a jump-diffusion process," *Z. Wahrscheinlichtkeitsteorie*, 53, 207-219 (1980).
9. H. J. Kushner and Hai Huang, "Averaging methods for the asymptotic analysis of learning and adaptive systems, with small adjustment rate," to appear *SIAM J. on Optimization and Control.*
10. G. C. Papanicolaou, D. Strook, and S. R. S. Varadham, "Martingale approach to some limit theorems," *Duke University Conference on Turbulence*, 1976.

11. H. J. Kushner, "Diffusion approximations to output processes of nonlinear systems with wide inputs," *IEEE Trans. on Information Theory*, IT-26, 715-725 (1980).

12. H. J. Kushner and Y. Bar-Ness, "Analysis of nonlinear systems with wide-band inputs," *IEEE Trans. on Automatic Control*, AC-25, 1072-1078 (1980).

13. H. J. Kushner and Hai Huang, "Diffusion approximations for the analysis of digital phase locked loops," LCDS Tech. Report 80-2, 1980, Brown University; also to appear.

14. H. J. Kushner and W. T. Y. Ju, "Diffusion approximations for nonlinear phase locked loop type systems with wide band inputs," LCDS Tech. Report 80-2, 1980, Brown University; to appear *J. Math. Anal. and Applic.*

15. H. J. Kushner, "Asymptotic distributions of solutions of ordinary differential equations with wide band noise inputs; approximate invariant measures," LCDS Tech. Report 81-2, January 1981, submitted.

16. P. Billingsley, *Convergence of Probability Measures* (New York: Wiley, 1968).

17. W. C. Lindsey, *Synchronization Systems in Communications and Control* (Englewood Cliffs, NJ: Prentice-Hall, 1972).

18. F. L. Gardner, *Phaselock Techniques* (New York: Wiley, 1979), 2nd edition.

19. H. J. Kushner and Hai Huang, "On the weak convergence of general stochastic difference equations to a diffusion," *SIAM J. on Applied Math.*, 40, 528-542 (1981).

jbf

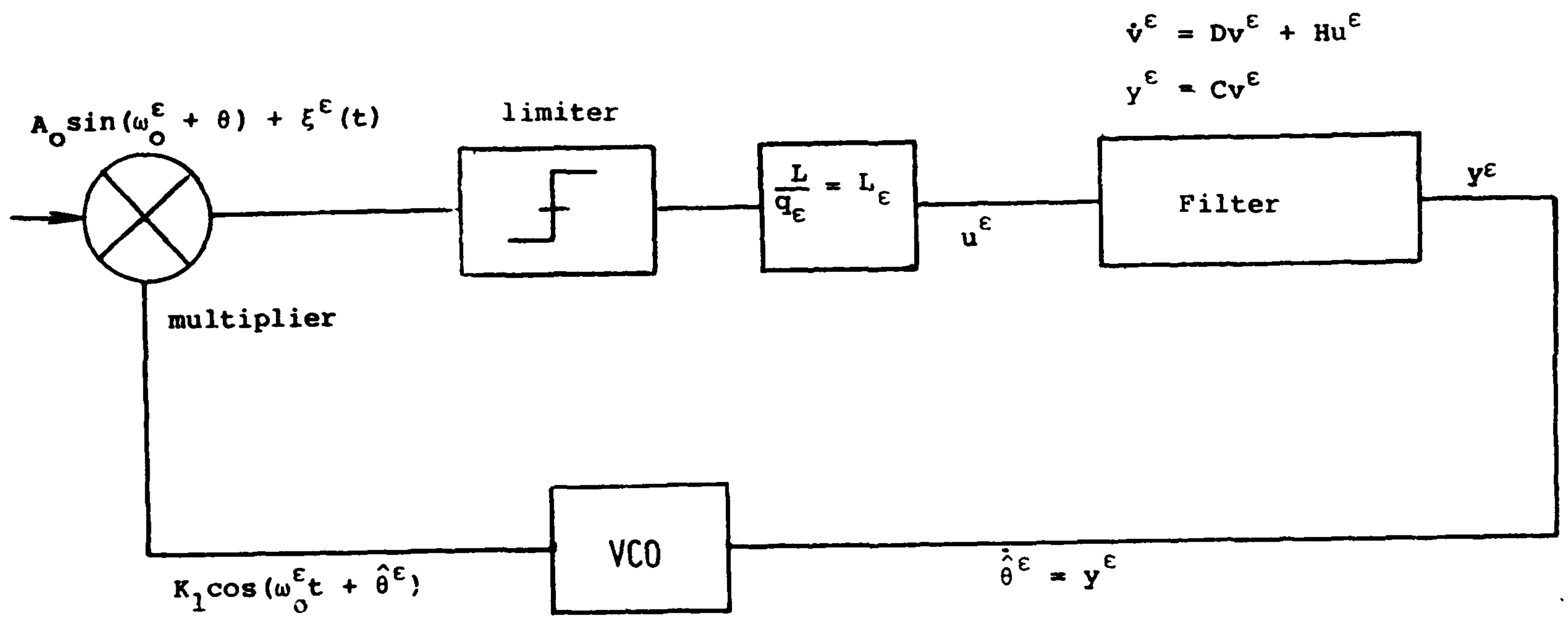

Figure 1

(refer to pp.11-20)